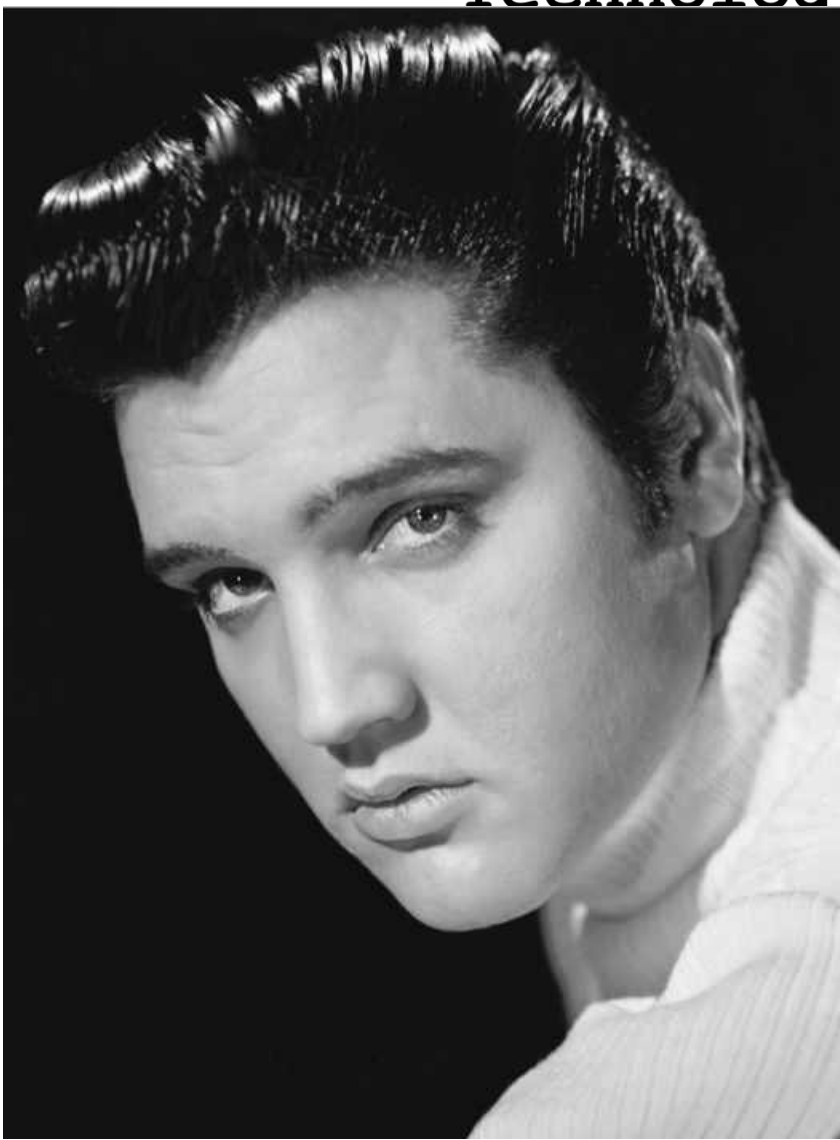




Elvis vs. M. Jackson: Who has More Albums?

Classification and Identification of Elements in Comparative Questions

Meriem Beloucif, Seid Muhie Yimam, Steffen Stahlhacke, and Chris Biemann



What is Comparative Question Answering (cQA)?

Classification Task

- The text classification task aims at identifying if a given question is comparative or not.

Difficulty	Correct	Sentences (blue for comparatives)
Easy	92%	what is the difference between a cappuccino and a latte?
Middle	62%	what can i do with a bachelors degree in history?
Easy	71%	should i buy or rent in California?
Easy	91%	what is the difference between burning and ripping?
Hard	43%	what are the differences between the those beams?
Easy	71%	can you please eli5 the difference between ham?
Middle	67%	what calculations can only be done by?
Hard	33%	why does everyone say hitting a pitch from a mlb?

- Comparative questions are generally treated the same as any open domain question, although they have a different taxonomy.
- Binary comparative questions are composed of two objects, comparison aspect, and attribute.

Comparative Questions Taxonomy

- We collected raw data from Yahoo! Answers, Reddit, and Quora.
- We studied the linguistic taxonomy for cQA with a better fit for popular web data.

Labeling Task

How to label **Comparative Objects** and **Aspects**:

Click on words to select a fitting label. Please mark objects that are compared to each other (entities, situations, people, real objects, companies, other comparable things, ...) and if available the aspect of the comparison. Please read the detailed Instructions and Examples!

Comparative Question

1. Why are **Samsung TV's** **brighter** than LG's ?

2. Are **Comparative-OBJ-1** **Comparative-OBJ-2**

3. what **Comparative-OBJ-3** **Comparative-OBJ-4** **Shared-OBJ** **Aspect** **No Label** and a dictator ?

Task:

Instructions

Examples and task definitions

How to use the Labeling Tool?

Search Definitions

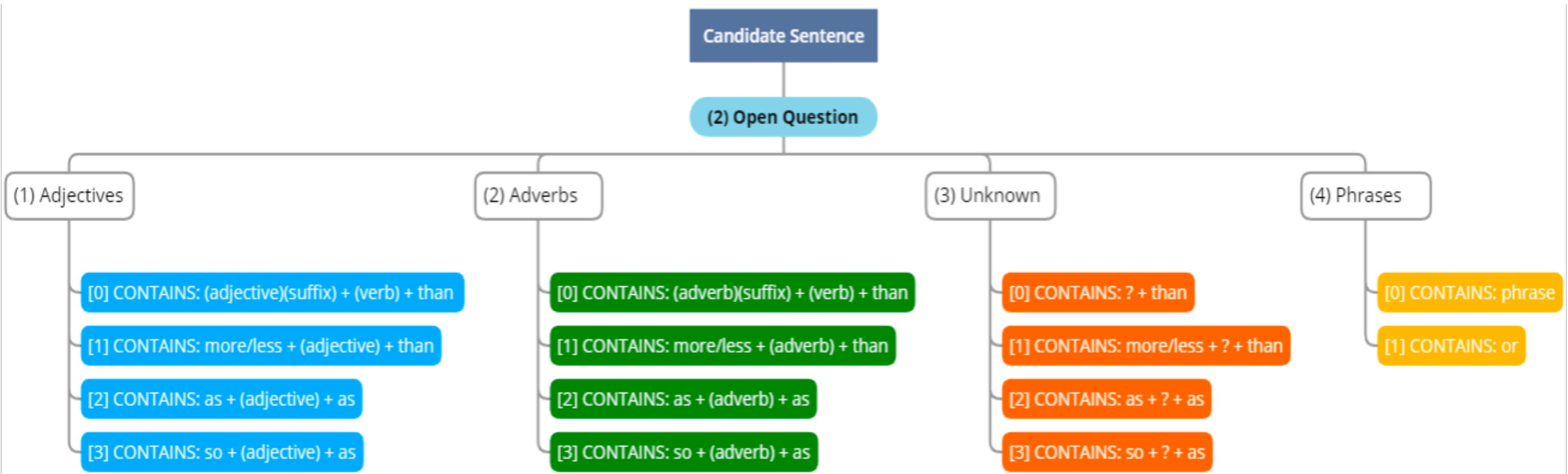
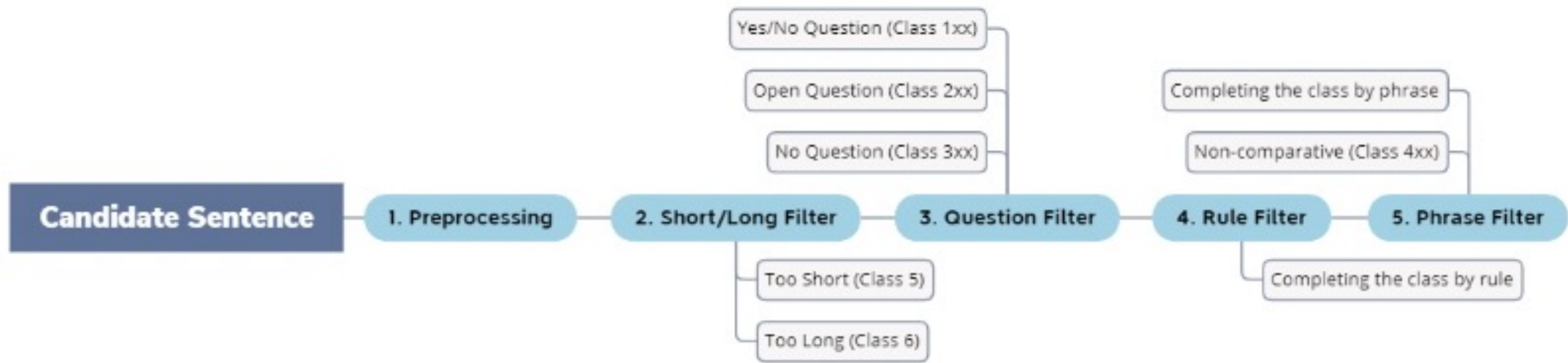
If you do not know the meaning of any terms in the texts below, write it here and search (opens in new window):

Search Definition Search

Results

Model	F1 Score Macro	F1 Score Comp.	F1 Score Non-comp
FLAIR LSTM with ALBERT large embedding	0.87	0.83	0.92
Human ceiling performance	0.89	0.89	0.90

Table 1. The performance of the best neural model for the classification task with the **ALBERT embeddings** is almost as good as human performance, but it beats the human when classifying not-comparative sentences.



Contact Information and Code

email: meriem.beloucif@lingfil.uu.se

<https://github.com/uhh-lt/Dataset-CompQA>

